Journal of Forestry Research (2013) 24(3): 495–502
DOI 10.1007/s11676-013-0380-7

ORIGINAL PAPER

Expressed sequence tags (ESTs) analysis of the ripening *Vitis amurensis* cv. Shuang Hong berry skins

JI Xiang-Nan • LI Feng • YANG Cheng-Jun • LI Bo
WANG Jun • ZHANG Wen

Received: 2012-02-15; Accepted: 2012-04-18

Abstract: *Vitis amurensis* is a valuable resource for wine production. Ripening of the grape berry is the key phase which determines the composition of wine. To better understand the gene expression that manifest in *V. amurensis* berry skins during the ripening, cDNA library of *V. amurensis* berry skins was constructed. A total of 935 high quality expressed sequence tags (ESTs) were obtained from the library. These ESTs represent 636 unigenes, including 108 contigs and 528 singletons. The EST analysis was performed and genes were assigned to functional categories according to their primary BLAST match. Of these 25.35% were involved with metabolism, 6.27% with cell rescue and defense, 6.84% energy, 11.68% protein synthesis, 18.8% protein activity regulation, 11.11% cell structure, 7.98% transport, 6.27% transcription and the remaining 5.7% were signal transduction. The generated ESTs were characterized by the gene ontology analysis and were categorized according to its cellular component, molecular function and biological process. In the cDNA library, some genes are relevant to the biosynthesis of anthocyanins, while some genes are related to grape berry maturation.

Fund project: This research was supported by the China Agriculture Research System (CARS-30).
The online version is available at http://www.springerlink.com

JI Xiang-Nan[☆] • WANG Jun
Center for Viticulture and Enology, College of Food Science & Nutritional Engineering, China Agricultural University, Beijing 100083, P.R. China

ZHANG Wen (✉)
College of Agriculture & Biotechnology, China Agricultural University, Beijing 100193, P.R. China.
E-mail: zhangwen@cau.edu.cn

LI Feng[☆], YANG Cheng-Jun, LI Bo,
College of Forestry, Northeast Forestry University, Harbin 150040, P.R. China
[☆]These authors contributed equally to this research.

Corresponding editor: Chai Ruihai

Key words: *Vitis amurensis*; expressed sequence tags (ESTs); cDNA library; unigene annotation; gene ontology

Introduction

V. amurensis (Amur grape), a famous wild-growing berry, is widely distributed in China, Korea, and Japan and its fruit has been used as the raw material for juice and wine in those three countries (Jeong et al. 2010). The wine made by *V. amurensis* is ruby, bright and has a unique flavor which favored by consumers. The ripening of grape berries is accompanied by a massive accumulation of soluble sugars, as well as the synthesis and accumulation of a wide range of phenolic compounds and aroma precursors. These processes play major roles in the quality of the berries and wine (Agasse et al. 2009). Besides, the skin is the site that anthocyanin biosynthesis occurs. During ripening phase, as the level of chlorophyll falls, the content of anthocyanins increases which bring distinctive color to colored berry skins and red wine. Therefore, it is important to investigate the genes related to the biosynthesis of anthocyanins in the *V. amurensis* ripen berry skins.

Analysis of expressed sequence tags (ESTs) is a rapid and effective method to identify novel genes or to investigate gene expression in different tissues, organs and plants. This technology can also identify genes that are involved in specific biological functions, especially for organisms whose genome sequences are not available (Peng et al. 2007). Gene functions are assigned to ESTs based on homology to known proteins from other species. Currently, about 50% of ESTs can be identified in this way (Ablett et al. 2000). During a specific development states, the abundance of ESTs of the cDNA library can be used to estimate the expression level of their corresponding genes (Fei et al. 2004; Ma et al. 2004). Analysis of the expression of large numbers of genes combined with knowledge of their functions allows us to perceive the overall picture of biological processes in different

cell types. These studies have been initiated by using primary BLAST homologues to divide ESTs matching known proteins into functional categories (Ablett et al. 2000). To 2011, more than 350,000 ESTs have been generated and analyzed for grapevine including both wine grape and table grape (Tillett et al. 2011).

As mentioned above, compared with *V. vinifera*, the molecular biological research of *V. amurensis* is limited. In order to analyze related gene expression in *V. amurensis* ripen berry skins, we constructed cDNA library of *V. amurensis* ripen berry skins and obtained 935 ESTs from it. These ESTs were assembled into 108 contigs and 528 singletons and were analyzed and characterized by BLAST alignments and gene ontology analysis.

Materials and methods

Plant materials

The cDNA library was derived from the berry skins of *V. amurensis* cv. Shuang Hong, the berries were collected 16 weeks after flowering from Institute of Special Wild Economic Animals and Plants, Chinese Academy of Agricultural Science. The berry skin and flesh were then separated, frozen immediately in liquid nitrogen and stored at -80℃. RNA was extracted from random sub-samples of -80℃ stored berry skin.

RNA extraction and cDNA library construction

The total RNA of *V. amurensis* ripen berry skins was isolated according to the protocol of Chang et al. (1993). The quality of RNA was analyzed by 1% agarose gel electrophoresis, Total RNA from *V. amurensis* ripen berry skins should appear as two bright bands (28S and 18S ribosomal RNA) at approximately 4.5 and 1.9 kb. The ratio of intensities of the 28S and 18S rRNA bands should be 1.5–2.5:1. The cDNA library was constructed as described by Creator™ SMART™ cDNA Library Construction Kit User Manual and the amount of starting material for cDNA synthesis is 0.8 g of total RNA.

DNA sequencing and data handling

Prior to sequencing, random colonies were checked for the presence of an insert by PCR using the M13 reverse primer and forward primer. The PCR products were separated by electrophoresis using 1.2% agarose gels. DNA sequencing was completed by Genomics Institute, Beijing, and P.R. China.

Raw single-pass sequence data demonstrating poor quality vector sequences or sequences less than 100 bp were removed. The remaining sequences were analyzed. Unigenes were obtained by Phrap assembly program. Unigenes were compared against the current non-redundant (nr) protein database at the National Center for Biotechnology Information (NCBI, Bethesda, MD, USA) using the BLASTX algorithm and the NCBI EST database using the BLASTN algorithm (Altschul et al. 1997).

Results and discussions

cDNA library characterization

The titer of primary cDNA library and amplified library was 1.001×10^{6} pfu·mL^{-1} and 2.542×10^{9} pfu·mL^{-1}, respectively. The cDNA clones selected randomly from the primary cDNA library were checked by PCR, and 94% were found to contain inserts with the size ranging from 0.3 to 2 kb. The mean insert size is 0.86 kb, which is shorter than those from grapevine cDNA libraries previously reported (e.g. 1.3 kb mean insert size in Terrier et al. 2001; 1.2 kb–1.5 kb mean insert size of six cDNA libraries of six different organs in Moser et al. 2005).

After removing vectors and low-quality sequences, we obtained a total of 935 high-quality EST sequences from the library. The average length of our EST sequences was approximately 416 bp (range 102 bp-611 bp), which is shorter than those obtained from grapevine cDNA libraries (e.g. 463 bases in Moser et al. 2005; 527 bases in da Silva et al. 2005; 622 bases in Peng et al. 2007). However, the majority of our ESTs (>88%) were longer than 100 bases. In conclusion, these results indicated that the cDNA library is a reliable resource for EST analysis.

Analysis of EST sequences

The sequences of the *V. amurensis* EST clones were submitted to NCBI as dbEST IDs 69043327-69044261 and GenBank accession Nos.GW665582-GW666516. After sequence comparison and analysis, we obtained 636 unique sequences, including 528 single-sequence and 108 multiple-sequence contigs with the frequency of occurrence of uniseqs ranging from 2 to 41. A contig is a group of cDNAs that share sequence identity and are considered to represent transcripts of the same gene. Sequencing only from the 5' terminus may leave some redundancy undetected. It is, therefore, possible that a gene may be represented by more than one contig (Sacadura and Saville 2003).

The EST sequences were compared to the NCBI Nr database using the BLASTX algorithms (Altschul et al. 1997). The result showed that a total of 522 unigenes (707 ESTs, 75.6%) exhibited high similarities ($E_{value}<10^{-5}$) to the genes sequences available in the Nr database, while the remaining114 unigenes (228 ESTs, 24.4%) showed little or no similarity. Further study needs to be carried out to determine whether these non-matching clones are new members of known gene families or novel genes. ESTs with similarity scores of $E_{value}<10^{-5}$ were grouped into 9 categories according to their functional annotation which involved in metabolism, cell resistance and defense, energy, protein synthesis, protein activity regulation, cell structure, transport and transcription (Fig. 1).

With the variation of developmental stages, and physiological and pathological states, different cells and organs of a certain organism have different expression types and abundance of a certain gene. The EST numbers of the same gene represent expression abundance of the gene in the specific organs (Audic and

Claverie 1997). In this cDNA library, approximately 95.9% of the ESTs belonged to sequence clusters with low redundancy (2-3 sequences per contig), while medium redundancy (4-9 sequences per contig) and high redundancy (≥10 sequence per contig) ESTs represented 3.3% and 0.7%, respectively (Fig. 2). This result implies that there is a considerable potential to discover novel sequences from our cDNA library (Jung et al. 2003).

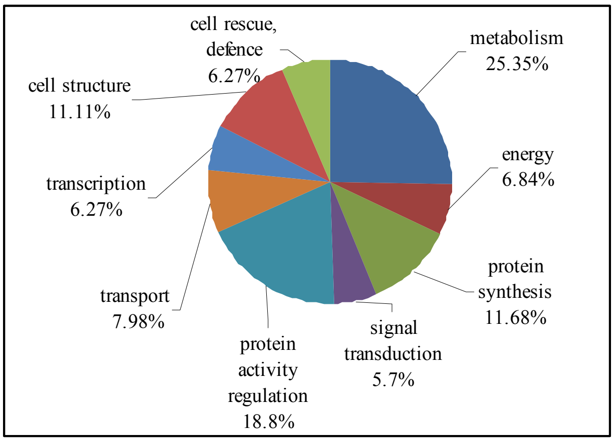

Fig. 1 Functional category of ESTs sequences in *V. amurensis* library

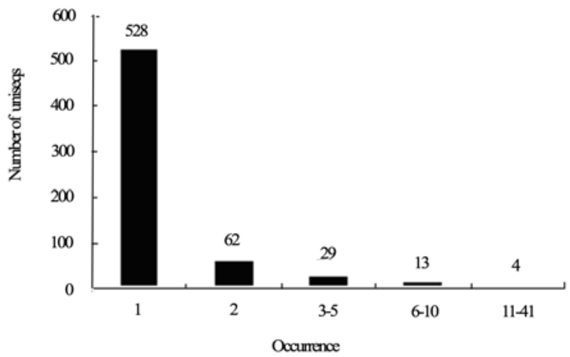

Fig. 2 Genes expressed frequency

A number of *V. amurensis* ESTs showed a great sequence similarity with ESTs from *V. vinifera* (171 clones, 32.8%, Table 1) *and Arabidopsis thaliana* (130 clones, 24.9%, Table 1). The following organisms are *Oryza sativa subsp. japonica, Glycine max, Nicotiana tabacum, Petunia hybrid, Triticum aestivum, Zea mays* according to priority of similarity.

By BLAST searching of NCBI databases, some ESTs are shown to share significant sequence similarity ($E_{value}<10^{-15}$) to a variety of known genes or gene families. In this study, there are 142 unigenes that had significant similarity to a variety of known genes or gene families (Table 2). For example, *V. amurensis* EST clone number 1-B09 was similar to chalcone synthase (97% identity), 04-D09 was similar to phenylalanine ammonia-lyase (95% identity), and conting132 was similar to ripening-related protein grip22 (91% identity).

In the cDNA library, the number of function genes, together with gene expression abundance, depends on the characteristic of the growth stage of a plant. The sampling time is the ripening of *V. amurensis* and the process involves the coordination of a large number of events. Some metabolic activities that occur prior to véraison, such as photosynthesis and organic acid accumulation, are either turned off at véraison or at least down regulated. Other processes, such as the accumulation of anthocyanins in berry skins, commence at véraison (Davies and Böttcher 2009).

Table 1 Organisms showing high sequence similarity with the *V. amurensis* ESTs

Organism	No. clones	Percentage（%）
Vitis vinifera	171	32.8%
Arabidopsis thaliana	130	24.9%
Oryza sativa subsp. *japonica*	22	4.2%
Glycine max	10	1.9%
Nicotiana tabacum	8	1.5%
Petunia hybrida	7	1.3%
Triticum aestivum	7	1.3%
Zea mays	7	1.3%
Solanum lycopersicum	6	1.1%
Solanum tuberosum	6	1.1%
Ricinus communis	4	0.8%
Catharanthus roseus	4	0.8%
Citrus sinensis	4	0.8%
Prunus avium	4	0.8%

The most frequent gene found in the cDNA library was ribosomal protein (28 clones, 5.38%). This result was expected because ribosomal protein genes are expressed ubiquitously at all stages of development (Kim et al. 2006). All protein synthesis needs to carry on the ribosome and the number of ribosomes influences the protein synthesis directly. This result indicated that the process of protein synthesis is active in the *V. amurensis* berry skins.

There are a number of different classes of flavonoids found in plants and anthocyanin is one of the most important pigments enduing grapes with colored skins, and giving red wine distinctive colors (Boss et al. 1996a). Anthocyanins are located in the skin cells, which present as a free, non-complex form inside the vacuoles (Ortega-Regules et al. 2008). The anthocyanin biosynthesis is vigorous at the ripening stage of grape. In the cDNA library we found some genes correlated with the biosynthesis of anthocyanins, such as chalcone synthase (*CHS*), flavonoid-3',5'-hydroxylase (*F3'5'H*), phenylalanine ammonia-lyase (*PAL*), glutathione *S*-transferases (*GST*) and anthocyanidin-3-glucosyltransferas.

PAL is the first enzyme involved in anthocyanin production, which catalyses the synthesis of cinnamic acid from phenylalanine (Boss and Davies 2009). In the growth process of red grape berries, the expression of *PAL* mainly occurred in the grape berry skins after flower 2 to 4 weeks and then decreased. The expression of *PAL* reached to climax during ripening (Boss et al. 1996a; Kobayashi et al. 2001). There are 15–20 members in grape *PAL* gene families (Sparvoli et al. 1994). The nucleotide sequence and the corresponding amino acid sequence of *PAL* are highly similar among different plants. In the cDNA library the *PAL* ESTs are shown to share significant sequence similarity

(E_{value}≤10^{-15}) to *Arabidopsis thaliana* and *Petroselinum crispum* (Table 2). This gene family possibly stems from an ancient gene, and the duplication of the gene together with the divergence of molecule probably product different function genes (Sparvoli et al. 1994).

Table 2. Unigenes have high similarity to a variety of known genes or gene families

Clone No.	Putative homologue	Accession No.	Organism	E-value	Score (bits)	Identity (%)
Contig101	Polygalacturonase inhibitor	A7PW81	*Vitis vinifera*	2.00E-27	120	100
Contig108	Agamous-like MADS-box protein AGL9 homolog	Q03489	*Petunia hybrida*	2.00E-75	281	95
Contig11	ADP-ribosylation factor 2	P51823	*Oryza sativa* subsp. *japonica*	3.00E-39	160	100
Contig113	Ubiquitin-conjugating enzyme E2-17 kDa	P35135	*Solanum lycopersicum*	7.00E-82	302	95
Contig119	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic	P25861	*Antirrhinum majus*	3.00E-52	204	90
Contig12	Glucan endo-1,3-beta-glucosidase	P52408	*Antirrhinum majus*	4.00E-26	116	76
Contig121	Chalcone synthase 1	Q9XJ58	*Citrus sinensis*	6.00E-39	159	80
Contig13	Flavonoid 3',5'-hydroxylase 2	P48419	*Petunia hybrida*	2.00E-74	278	80
Contig131	Glutathione *S*-transferase	Q96324	*Arabidopsis thaliana*	3.00E-69	261	58
Contig132	Ripening-related protein grip22	Q9M4H4	*Vitis vinifera*	1.00E-118	426	91
Contig133	Uncharacterized mitochondrial protein AtMg00030	P93276	*Arabidopsis thaliana*	4.00E-28	124	96
Contig134	Peptidyl-prolyl *cis-trans* isomerase	Q39613	*Catharanthus roseus*	2.00E-36	151	91
Contig15	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic	P26518	*Magnolia liliiflora*	2.00E-29	127	87
Contig16	Histone H4	Q76H85	*Silene latifolia*	2.00E-39	157	97
Contig17	ATP synthase subunit epsilon, mitochondrial	Q96253	*Arabidopsis thaliana*	9.00E-27	119	87
Contig19	Flavonoid 3',5'-hydroxylase 2	P48419	*Petunia hybrida*	6.00E-18	89.4	80
Contig29	GTP-binding nuclear protein Ran-B1	P41919	*Nicotiana tabacum*	2.00E-66	252	97
Contig30	Ubiquitin-like protein 5	Q9FGZ9	*Arabidopsis thaliana*	2.00E-35	148	95
Contig31	Alcohol dehydrogenase 1	P25141	*Petunia hybrida*	1.00E-144	509	81
Contig32	Vacuolar-processing enzyme	P49043	*Citrus sinensis*	1.00E-61	234	82
Contig34	Peptidyl-prolyl isomerase FKBP12	O04287	*Vicia faba*	3.00E-53	206	83
Contig38	Polygalacturonase	P35336	*Actinidia deliciosa*	2.00E-39	161	65
Contig40	14-3-3-like protein B (Fragment)	Q96451	*Glycine max*	1.00E-44	178	91
Contig45	Ubiquitin	P69326	*Triticum aestivum*	4.00E-36	150	100
Contig47	Probable aquaporin PIP1-2	Q7XSQ9	*Oryza sativa* subsp. *japonica*	2.00E-61	234	91
Contig5	Histone H4	Q76H85	*Silene latifolia*	5.00E-40	162	100
Contig54	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic	P26518	*Magnolia liliiflora*	2.00E-63	241	81
Contig55	Phenylalanine ammonia-lyase	P35510	*Arabidopsis thaliana*	3.00E-84	311	78
Contig57	Ubiquitin	P69326	*Triticum aestivum*	3.00E-36	150	100
Contig58	40S ribosomal protein S26-3	Q9LYK9	*Arabidopsis thaliana*	6.00E-16	83.2	86
Contig6	ADP-ribosylation factor 2	P51823	*Oryza sativa* subsp. *japonica*	1.00E-61	235	99
Contig60	Pyruvate kinase, cytosolic isozyme	Q42806	*Glycine max*	3.00E-61	235	90
Contig67	Ubiquitin-like protein SMT3	P55852	*Arabidopsis thaliana*	3.00E-42	171	88
Contig71	Ras-related protein ARA-1	P19892	*Arabidopsis thaliana*	6.00E-79	293	84
Contig73	30S ribosomal protein S7, chloroplastic	Q0ZIV9	*Vitis vinifera*	4.00E-60	229	100
Contig76	Transcription elongation factor 1 homolog	Q8LHP0	*Oryza sativa* subsp. *japonica*	2.00E-31	135	88
Contig77	Isocitrate dehydrogenase [NADP], chloroplastic (Fragment)	Q40345	*Medicago sativa*	1.00E-120	432	89
Contig78	Peptide methionine sulfoxide reductase (Fragment)	P54153	*Solanum lycopersicum*	4.00E-80	297	80
Contig8	Uncharacterized mitochondrial protein AtMg00030	P93276	*Arabidopsis thaliana*	9.00E-29	125	98
Contig80	ATP synthase subunit epsilon, mitochondrial	Q96253	*Arabidopsis thaliana*	1.00E-26	119	85
Contig82	Probable glutathione *S*-transferase	Q03666	*Nicotiana tabacum*	1.00E-82	306	72
Contig86	60S ribosomal protein L27a-3	P49637	*Arabidopsis thaliana*	1.00E-72	271	86
Contig88	60S ribosomal protein L35a-3	Q9C912	*Arabidopsis thaliana*	4.00E-56	217	91
Contig91	50S ribosomal protein L2, chloroplastic	Q0ZIX7	*Vitis vinifera*	2.00E-72	273	99
Contig93	Fructose-bisphosphate aldolase cytoplasmic isozyme	P17784	*Oryza sativa* subsp. *japonica*	1.00E-57	222	90
Contig96	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic	P25861	*Antirrhinum majus*	5.00E-58	223	88
Contig97	Glutamine synthetase cytosolic isozyme 1	P51118	*Vitis vinifera*	2.00E-70	265	99
0001_A02	14-3-3 protein 4	P42652	*Solanum lycopersicum*	4.00E-65	246	93
0001_B09	Chalcone synthase	P51090	*Vitis vinifera*	2.00E-79	294	97
0001_D04	Isopentenyl-diphosphate Delta-isomerase I	O48964	*Camptotheca acuminata*	2.00E-44	177	91

Continue Table 2

Clone No.	Putative homologue	Accession No.	Organism	E-value	Score (bits)	Identity (%)
0001_G11	Monodehydroascorbate reductase	Q40977	*Pisum sativum*	1.00E-49	194	80
0001_H12	Histone H2A	Q9M531	*Euphorbia esula*	7.00E-17	85.9	100
0002_B01	Ras-related protein Rab7	O24461	*Prunus armeniaca*	1.00E-41	169	95
0002_C01	24-methylenesterol *C*-methyltransferase 2	Q39227	*Arabidopsis thaliana*	2.00E-85	314	89
0002_C03	Elongation factor 1-alpha	P25698	*Glycine max*	3.00E-26	117	94
0002_D12	Shaggy-related protein kinase zeta	Q39010	*Arabidopsis thaliana*	1.00E-86	318	91
0002_F09	T-complex protein 1 subunit epsilon	P54411	*Avena sativa*	3.00E-18	90.1	89
0002_G05	Acyl-CoA-binding protein	O04066	*Ricinus communis*	8.00E-18	89	87
0002_G06	Uncharacterized FAM18-like protein At1g09330	Q8LEK2	*Arabidopsis thaliana*	2.00E-76	284	80
0003_D10	Probable aquaporin PIP2-5	Q9SV31	*Arabidopsis thaliana*	2.00E-21	101	83
0003_E11	Enolase	Q42971	*Oryza sativa* subsp. *japonica*	6.00E-82	303	86
0004_A05	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha	Q41140	*Ricinus communis*	2.00E-80	298	80
0004_B02	Eukaryotic peptide chain release factor subunit 1-3	P35614	*Arabidopsis thaliana*	5.00E-46	182	85
0004_B03	40S ribosomal protein S12	Q9XHS0	*Hordeum vulgare*	3.00E-40	164	81
0004_B08	F-box protein GID2	Q9STX3	*Arabidopsis thaliana*	2.00E-27	122	80
0004_C11	60S ribosomal protein L29-1	Q9M7X7	*Arabidopsis thaliana*	5.00E-25	112	83
0004_C12	Cytochrome P450 77A1 (Fragment)	P37123	*Solanum melongena*	1.00E-22	104	82
0004_D09	Phenylalanine ammonia-lyase 1	P24481	*Petroselinum crispum*	3.00E-85	313	95
0004_D10	60S ribosomal protein L3	P35684	*Oryza sativa* subsp. *japonica*	4.00E-17	86.7	85
0004_D12	Eukaryotic translation initiation factor 1A	P56331	*Onobrychis viciifolia*	7.00E-51	199	95
0004_F12	3-ketoacyl-CoA thiolase 2, peroxisomal	Q56WD9	*Arabidopsis thaliana*	2.00E-38	157	90
0004_G02	Ferritin-3, chloroplastic	Q948P6	*Glycine max*	5.00E-37	120	84
0004_G10	60S ribosomal protein L7-3	P60039	*Arabidopsis thaliana*	7.00E-62	235	83
0005_B03	SUMO-conjugating enzyme UBC9	P35132	*Arabidopsis thaliana*	1.00E-70	265	98
0005_B11	*S*-adenosylmethionine synthetase 4	A7PRJ6	*Vitis vinifera*	1.00E-93	341	92
0005_C02	Ubiquitin-conjugating enzyme E2 2	P35130	*Medicago sativa*	2.00E-85	314	97
0005_D01	Eukaryotic translation initiation factor 4E type 3	Q9FK59 I	*Arabidopsis thaliana*	3.00E-90	330	81
0005_E06	ADP-ribosylation factor 2	P51823	*Oryza sativa* subsp. *japonica*	2.00E-76	285	97
0006_A11	DNA-damage-repair/toleration protein DRT102	Q05212	*Arabidopsis thaliana*	4.00E-34	143	84
0006_B01	Hydroxymethylglutaryl-CoA synthase	P54873	*Arabidopsis thaliana*	8.00E-88	322	83
0006_B02	Methionine aminopeptidase 2B	Q56Y85	*Arabidopsis thaliana*	5.00E-25	113	94
0006_B04	Protein translation factor SUI1 homolog	Q0D5W6 J	*Oryza sativa* subsp. *japonica*	8.00E-55	213	93
0006_B06	Stress-related protein	Q9SW70	*Vitis riparia*	1.00E-85	314	96
0006_B11	Elongation factor 1-alpha	P25698	*Glycine max*	2.00E-63	241	95
0006_C06	Shaggy-related protein kinase eta	Q39011	*Arabidopsis thaliana*	7.00E-26	116	80
0006_C10	60S ribosomal protein L23	P49690	*Arabidopsis thaliana*	1.00E-56	218	98
0006_C12	60S ribosomal protein L24	Q9FUL4	*Prunus avium*	9.00E-48	188	96
0006_D02	Ubiquitin-conjugating enzyme E2-17 kDa	P35135	*Solanum lycopersicum*	4.00E-83	306	97
0006_D12	60S ribosomal protein L32-1	P49211	*Arabidopsis thaliana*	6.00E-20	95.9	92
0006_E01	Cell division cycle protein 48 homolog	P54774	*Glycine max*	3.00E-21	100	85
0006_E08	Hemoglobin-2	P23244	*Casuarina glauca*	1.00E-68	258	84
0006_F01	Caffeic acid 3-*O*-methyltransferase	Q43609	*Prunus dulcis*	2.00E-93	341	83
0006_H03	Uncharacterized protein At5g10860, mitochondrial	Q9LEV3	*Arabidopsis thaliana*	3.00E-84	310	82
0006_H07	Anthocyanidin 3-*O*-glucosyltransferase (Fragment)	P51094	*Vitis vinifera*	4.00E-30	129	100
0007_A05	FAM10 family protein At4g22670	Q93YR3	*Arabidopsis thaliana*	9.00E-42	169	88
0007_B03	GTP-binding protein SAR1A	O04834	*Arabidopsis thaliana*	3.00E-31	134	93
0007_C10	Cytochrome c oxidase subunit 1	P20681	*Podo ora anserina*	7.00E-93	339	90
0007_D06	Ubiquitin	P69326	*Triticum aestivum*	4.00E-36	150	100
0007_E10	A aragine synthetase [glutamine-hydrolyzing]	Q43011	*Oryza sativa* subsp. *japonica*	4.00E-52	203	80
0007_H01	Thiazole biosynthetic enzyme, chloroplastic	O23787	*Citrus sinensis*	4.00E-19	93.6	84
0007_H04	Thiazole biosynthetic enzyme, chloroplastic	O23787	*Citrus sinensis*	1.00E-74	278	85
007-G06	Pectinesterase	Q9LVQ0	*Arabidopsis thaliana*	9.00E-61	231	76
0008_A05	GTP-binding nuclear protein Ran1B (Fragment)	P54766	*Lotus japonicus*	2.00E-36	150	80
0008_A07	Actin-depolymerizing factor	Q8SAG3	*Vitis vinifera*	5.00E-58	223	91
0008_A10	BURP domain-containing protein 3	Q942D4	*Oryza sativa* subsp. *japonica*	3.00E-19	93.6	80

Continue Table 2

Clone No.	Putative homologue	Accession No.	Organism	E-value	Score (bits)	Identity (%)
0008_A11	Probable cellulose synthase A catalytic subunit 8 [UDP-forming]	Q84ZN6	*Oryza sativa* subsp. *japonica*	2.00E-78	291	88
0008_B06	VAMP-like protein YKT61	Q9ZRD6	*Ricinus communis*	7.00E-79	292	86
0008_E03	Vesicle-associated membrane protein 722	P47192	*Arabidopsis thaliana*	3.00E-87	320	80
0008_E08	Pyruvate kinase, cytosolic isozyme	Q42806	*Glycine max*	4.00E-57	219	91
0008_F02	Leucoanthocyanidin dioxygenase	P51091	*Malus domestica*	2.00E-70	263	81
0008_H05	UPF0497 membrane protein 7	A7PA04	*Vitis vinifera*	9.00E-46	181	100
0008_H07	Elongation factor 1-gamma	Q9FUM1	*Prunus avium*	1.00E-77	288	89
0009_A02	Cysteine proteinase (Fragment)	P05993	*Carica papaya*	9.00E-24	108	90
0009_A08	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	Q42662	*Solenostemon scutellarioides*	7.00E-95	345	96
0009_C04	50S ribosomal protein L33, chloroplastic	Q0ZIZ9	*Vitis vinifera*	4.00E-32	137	95
0009_D02	Catalase isozyme 1 (Fragment)	P49315	*Nicotiana plumbaginifolia*	2.00E-60	231	88
0009_D11	Light-induced protein, chloroplastic	P80471	*Solanum tuberosum*	6.00E-44	176	84
0009_E06	Ribosomal protein S19, mitochondrial	P27527	*Petunia hybrida*	9.00E-36	149	80
0009_F02	UTP--glucose-1-pho hate uridylyltransferase	O64459	*Pyrus pyrifolia*	2.00E-91	334	89
0009_F06	Putative multidrug resistance-associated protein 15	Q7FB56	*Arabidopsis thaliana*	1.00E-53	208	86
0009_G10	GTP-binding protein SAR1A	O04834	*Arabidopsis thaliana*	2.00E-53	171	95
0009_H11	F-box protein ORE9	Q9SIM9	*Arabidopsis thaliana*	5.00E-36	149	84
0010_A04	Heat shock protein 81-1	Q0J4P2	*Oryza sativa* subsp. *japonica*	3.00E-55	213	94
0010_A12	Proteasome subunit beta type-6	Q8LD27	*Arabidopsis thaliana*	9.00E-30	129	86
0010_C04	Pyropho hate-energized vacuolar membrane proton pump	P21616	*Phaseolus aureus*	3.00E-28	123	91
0010_C11	Translationally-controlled tumor protein homolog	Q43847	*Solanum tuberosum*	6.00E-21	100	81
0010_D08	Superoxide dismutase [Cu-Zn] 4AP	P23346	*Zea mays*	2.00E-15	80.9	94
0010_E07	Uncharacterized protein At5g01610	Q9M015	*Arabidopsis thaliana*	1.00E-45	181	82
0010_E09	Obtusifoliol 14-alpha demethylase (Fragment)	P93596	*Triticum aestivum*	3.00E-65	246	87
0010_F10	Alpha-soluble NSF attachment protein	P93798	*Vitis vinifera*	6.00E-36	149	91
0010_G01	60S ribosomal protein L17-2	P51413	*Arabidopsis thaliana*	6.00E-72	269	88
0011_C02	Multidrug resistance-associated protein 6	Q8LGU1	*Arabidopsis thaliana*	2.00E-43	174	88
0011_C06	Cell division protease ftsH homolog 1, chloroplastic	Q5Z974	*Oryza sativa* subsp. *japonica*	1.00E-100	363	96
0011_D03	Translationally-controlled tumor protein homolog	Q5J907	*Elaeis guineensis var. tenera*	6.00E-33	139	89
0011_F01	60S ribosomal protein L37a	Q9XHE4	*Gossypium hirsutum*	5.00E-47	186	98
0011_F07	Obtusifoliol 14-alpha demethylase (Fragment)	P93596	*Triticum aestivum*	5.00E-50	196	90
0011_G08	60S ribosomal protein L11-1	P42795	*Arabidopsis thaliana*	7.00E-16	82.4	90
0011_G11	Uncharacterized protein At2g23090	O64818	*Arabidopsis thaliana*	1.00E-30	132	86
0011_G12	Calmodulin	P62201	*Lilium longiflorum*	5.00E-72	269	99
0011_H05	Aquaporin PIP1-3	Q08733	*Arabidopsis thaliana*	1.00E-63	241	88
0012_C03	Expansin-A4	Q0DHB7	*Oryza sativa* subsp. *japonica*	2.00E-76	284	82
0012_G06	Calreticulin	P93508	*Ricinus communis*	6.00E-62	236	82

CHS is the first enzyme in the flavonoid biosynthetic pathway and chalcone is the first flavonoid produced (Boss and Davies 2009). In the growth process of red grape berries, the expression of *CHS* mainly occurred in the grape berry skins after flower 2 to 4 weeks and then decreased. The expression of *PAL* reached to climax during ripening (Boss et al. 1996a; Kobayashi et al. 2001; Waters et al. 2005). The *CHS* of grape was coded by multiple genes families and there are 3-4 members involved. The *CHS* EST involved in the cDNA library is shown to share significant sequence similarity (E_{value}≤10^{-15}) to *Arabidopsis thaliana* (Table 2).

F3’5’H belongs to the gene families of cytochrome *P450*. During the berry ripening, the transcription abundance of *F3’5’H* is closely related to hydroxylation level of anthocyanins. The proposal is that the transcription level of *F3’5’H* determines the ratio of cyanidin anthocyanins and delphinidin anthocyanins, and then influences berry skins color (Bogs et al. 2006; Jeong et al. 2006).

GST has a range of functions including the transport of anthocyanins into the vacuole. The anthocyanins is one of *GSTs* substrates in plant, such as *Bz2* (type III GST) of *Zea mays* and *An9* (type I plant *GST*) of *Petunia hybrid* regulated by conserved transcriptional activators during anthocyanins biosynthetic pathway. Coded enzyme protein may urge the anthocyanins glutathione and deposits in the vacuole (Alfenito et al. 1998; Wang and Liu 2008). Therefore *GST* can be considered the final enzyme in the anthocyanins biosynthetic pathway (Alfenito et al. 1998; Terrier et al. 2005; Ageorges et al. 2006).

Anthocyanidin-3-glucosyltransferase is the only and key gene whose expression is correlated with anthocyanin biosynthesis (Boss et al. 1996a; Boss et al. 1996b), the expression of which determines whether or not anthocyanins are synthesized (Waters et al. 2005).

There are some genes related to grape berry maturation in this

cDNA library, such as polygalacturonase (*PG*), ripening-related protein grip22, pectinesterase (*PE*), and glucosan interior contact-1, 3-beta-grape glycosidase and so on. *PG* is the cell wall hydrolytic enzymes that expresses specifically during berry ripening and participates in pectin dissolution, which is also the major enzyme of berry soften processes. Ripening-related protein grip22 gene has been proved to give rise to ripening of grape berries and may be related to thematic proteins, which is a class of proteins involved in disease resistance (Boss et al. 1996b). The *PE* increase solubility of the pectin in water in order to favor *PG* function.

The Gene Ontology (GO) analysis has been widely used to characterize gene function annotation and classification (Tanguy et al. 2008; Uno et al. 2008). In this study, the generated ESTs were categorized using GO terms as shown in Table 3, which provide a structured vocabulary to describe a sequence according to its cellular component, molecular function and biological process. Most ESTs appeared to be related to physiological processes. The result showed that the ripening process of *V. amurensis* is a complex physiological process.

Table 3 Gene Ontology category of ESTs in *V. amurensis* library

Gene Ontology term	No. clones sequenced
Cell component	
Cell	93
protein complex	37
organelle	41
extracellular region	2
Molecular Function	
antioxidant activity	2
catalytic activity	143
binding	127
structural molecule activity	24
transcription regulator activity	4
transporter activity	36
enzyme regulator activity	2
nutrient reservoir activity	3
signal transducer activity	4
translation regulator activity	6
Biological process	
cellular process	156
physiological process	186
regulation of biological process	8
response to stimulus	13
reproduction	1

Conclusion

In this study we described the construction of a cDNA library from *V. amurensis* and obtained 935 high quality EST sequences. These genes showed diverse functions involved in cellular component, molecular function and biological processes according to GO annotation. Our results suggested that the ripening of *V. amurensis* is a complex process containing multiple physiological and metabolic pathways. Our ESTs analysis can be helpful for further molecular mechanisms and genetic studies of *V. amurensis*.

References

Ablett E, Seaton G, Scott K, Shelton D, Graham MW, Baverstock P, Lee LS, Henry R. 2000. Analysis of grape ESTs: global gene expression patterns in leaf and berry. *Plant Science,* **159**: 87–95.

Agasse A, Vignault C, Kappel C, Conde C, Gerós H, Delrot S. 2009. Sugar transport & sugar sensing in grape. In: Roubelakis-Angelakis, Kalliopi A. (ed.), *Grapevine Molecular Physiology & Biotechnology.* Berlin: Springer-Verlag, pp. 105–139.

Ageorges A, Fernandez L, Vialet S, Merdinoglu D, Terrier N, Romieu C. 2006. Four specific isogenes of the anthocyanin metabolic pathway are systematically co-expressed with the red colour of grape berries. *Plant Science,* **170**: 372–383.

Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R, Walbot V. 1998. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione *S*-transferases. *The Plant Cell,* **10**: 1135–1149.

Altschul SF, Madden TL, Schäffe, AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* **25**: 3389–3402.

Audic S, Claverie J-M. 1997. The significance of digital gene expression profiles. *Genome Research,* **7**: 986–995.

Bogs J. 2005. Identification of the flavonoid hydroxylases from grapevine and their regulation during fruit development. *Plant Physiology,* **140**: 279–291.

Boss PK, Davies C. 2009. Molecular biology of anthocyanin accumulation in grape berries. In: Roubelakis-Angelakis, Kalliopi A. (ed.), *Grapevine Molecular Physiology & Biotechnology.* Berlin: Springer-Verlag, pp. 263–292.

Boss PK, Davies C, Robinson SP. 1996a. Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L. cv. Shiraz grape berries and the implications for pathway regulation. *Plant Physiology,* **111**: 1059–1066.

Boss PK, Davies C, Robinson SP. 1996b. Expession of anthocyanin biosynthesis pathway genes in red and white grapes. *Plant Molecular Biology,* **32**: 565–569.

Chang S, Puyrear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Moecularl Biology Reporter,* **11**: 113–116.

Davies C, Böttcher C. 2009. Hormonal control of grape berry ripening. In: Roubelakis-Angelakis, Kalliopi A. (ed.), *Grapevine Molecular Physiology & Biotechnology*. Berlin: Springer-Verlag, pp: 229–261.

da Silva FG, Iandolino A, Al-Kayal F, Bohlmann MC, Cushman MA, Lim H, Ergul A, Figueroa R., Kabuloglu EK, Osborne C, Rowe J, Tattersall E, Leslie A, Xu J, Baek JM, Cramer GR, Cushman JC, Cook DR. 2005. Characterizing the grape transcriptome. Analysis of expressed sequence tags from multiple *Vitis* species and development of a compendium of gene expression during berry development. *Plant Physiology,* **139**: 574–597.

Fei Z, Tang X, Alba RM, White JA, Ronning CM, Martin GB, Tanksley SD, Giovannoni JJ. 2004. Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. *The Plant Journal,* **40**: 47–59.

Jeong HY, Kim JY, Lee HK, Ha DT, Song K-S, Bae K, Seong YH. 2010. Leaf and stem of *Vitis amurensis* and its active components protect against

amyloid β protein (25–35)-induced neurotoxicity. *Archives of Pharmacal Research,* **33**: 1655–1664.

Jeong ST, Goto-Yamamoto N, Hashizume K, Esaka M. 2006. Expression of the flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase genes and flavonoid composition in grape (*Vitis vinifera*). *Plant Science,* **170**: 61–69.

Jung JD, Park HW, Hahn Y, Hur CG, In DS, Chung HJ, Liu JR, Choi DW. 2003. Discovery of genes for ginsenoside biosynthesis by analysis of ginseng expressed sequence tags. *Plant Cell Reports,* **22**: 224–230.

Kobayashi S, Ishimaru, M, Ding CK, Yakushiji H, Goto N. 2001. Comparison of UDP-glucose: flavonoid 3-*O*-glucosyltransferase (UFGT) gene sequences between white grapes (*Vitis vinifera*) and their sports with red skin. *Plant Science,* **160**: 543–550.

Ma HM, Schulze S, Lee S, Yang M, Mirkov E, Irvine J, Moore P, Paterson A. 2004. An EST survey of the sugarcane transcriptome. *Theoretical and Applied Genetics,* **108**: 851–863.

Moser C, Segala C, Fontana P, Salakhudtinov I, Gatto P, Pindo M, Zyprian E, Toepfer R, Grando MS, Velasco R. 2005. Comparative analysis of expressed sequence tags from different organs of *Vitis vinifera* L. *Functional & Integrative Genomics,* **5**: 208–217.

Ortega-Regules A, Ros-García JM, Bautista-Ortín AB, López-Roca JM, Gómez-Plaza E. 2007. Differences in morphology and composition of skin and pulp cell walls from grapes (*Vitis vinifera* L.): technological implications. *European Food Research and Technology,* **227**: 223–231.

Peng FY, Reid KE, Liao N, Schlosser J, Lijavetzky D, Holt R, Martínez Zapater JM, Jones S, Marra M, Bohlmann J. 2007. Generation of ESTs in *Vitis vinifera* wine grape (Cabernet Sauvignon) and table grape (Muscat Hamburg) and discovery of new candidate genes with potential roles in berry development. *Gene,* **402**: 40–50.

Sacadura NT, Saville BJ. 2003. Gene expression and EST analyses of *Ustilago maydis* germinating teliospores. *Fungal Genetics and Biology,* **40**: 47–64.

Sparvoli F, Martin C, Scienza A, Gavazzi G, Tonelli C. 1994. Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (*Vitis vinifera* L.). *Plant Molecular Biology,* **24**: 743–755.

Tanguy A, Bierne N, Saavedra C, Pina B, Bachère E, Kube M, Bazin E, Bonhomme F, Boudry P, Boulo V, Boutet I, Cancela L, Dossat C, Favrel P, Huvet A, Jarque S, Jollivet D, Klages S, Lapègue S, Leite R, Moal J, Moraga D, Reinhardt R, Samain J-F, Zouros E, Canario A. 2008. Increasing genomic information in bivalves through new EST collections in four species: Development of new genetic markers for environmental studies and genome evolution. *Gene,* **408**: 27–36.

Terrier N, Ageorges A, Abbal P, Romieu C. 2001. Generation of ESTs from grape berry at various developmental stages. *Journal of Plant Physiology,* **158**: 1575–1583.

Terrier N, Glissant D, Grimplet J, Barrieu F, Abbal P, Couture C, Ageorges A, Atanassova R, Léon C, Renaudin J-P, Dédaldéchamp F, Romieu C, Delrot S, Hamdi S. 2005. Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (*Vitis vinifera* L.) development. *Planta,* **222**: 832–847.

Tillett RL, Ergül A, Albion RL, Schlauch KA, Cramer GR, Cushman JC. 2011. Identification of tissue-specific, abiotic stress-responsive gene expression patterns in wine grape (*Vitis vinifera* L.) based on curation and mining of large-scale EST data sets. *BMC Plant Biology,* **11**: 86.

Uno Y, Suzuki Y, Wakaguri H, Sakamoto Y, Sano H, Osada N, Hashimoto K, Sugano S, Inoue I. 2008. Expressed sequence tags from cynomolgus monkey (*Macaca fascicularis*) liver: A systematic identification of drug-metabolizing enzymes. *FEBS Letters,* **582**: 351–358.

Wang J, Liu CP. 2008. Anthocyanin biosynthesis in grapevine. *Plant Physiology Communications*, **44**: 363–377. (In Chinese)

Waters DLE, Holton TA, Ablett EM, Lee LS, Henry RJ. 2004. cDNA microarray analysis of developing grape (*Vitis vinifera* cv. Shiraz) berry skin. *Functional & Integrative Genomics,* **5**: 40–58.